The Journal of Infectious Diseases

EDITORIAL COMMENTARY

# Causal Inference for Observational Studies

**David Kaplan**

Case Western Reserve University, Cleveland, Ohio

**(See the Major Article by Nguyen et al on pages 10–8.)**

"Correlation does not prove causation." This tired truism has been used to bolster experimental studies in preference to observational investigations for years without recognition that experimentation introduces significantly more investigator bias. Of course, correlation does not prove causation, but it does not refute it either.

Experimental clinical studies are strictly regulated in the United States by the Food and Drug Administration. Consequently, most clinical studies are observational, and statisticians have developed tools for addressing causality in observational studies which have propelled the development of a satellite discipline of statistics, causal inference [1].

In this issue of *The Journal of Infectious Diseases*, Nguyen et al describe a study that relies on causal inference [2]. These authors used a causal inference technique, propensity score matching, to ascertain that treatment of patients chronically infected with hepatitis B virus (HBV) and treated with tenofovir disoproxil significantly decreased the incidence of hepatocellular carcinoma. The comparison group included infected patients who were not treated.

Propensity score matching is a powerful technology that allows for causal inference in this type of observational study [3]. It is becoming popular for clinical research since, for a variety of reasons, randomized clinical trials are not always feasible. The propensity score is the probability that a patient has been assigned to the treatment or control group, conditioned on a set of covariates. It is calculated by logistic regression including the covariates as the independent variables and predicting patient stratification as the dependent variable. The selection of the covariates to include in the model represents an important facet of the analysis since it determines how the patients are matched.

After calculating the propensity scores for each subject, the investigators use these scores to match the patients in the treated group to the patients in the control group. In this way, a randomized trial is approximated on the basis of covariates in the investigators' model. Nguyen et al used caliper matching to select the control group, which refers to matching within a given proportion of the standard deviation of the propensity score. The original caliper used was 0.25 standard deviations [4], but the appropriateness of this variance was not subjected to critical analysis. A tighter caliper has been shown to provide reduced bias and closer matches, but it comes at the cost of decreasing the probability of finding matches [5]. In the study by Nguyen et al, a caliper of 0.2 standard deviations was used, and all treated subjects were successfully matched.

An analysis with propensity score matching requires a group of untreated subjects to match with the treated patients. To match most treated subjects, investigators have usually required 2–20 times more untreated subjects. In the study by Nguyen et al, the untreated group included 6140 subjects to match with the 774 patients treated with tenofovir disoproxil. The ratio of 8:1 was sufficient to allow for matching of all treated patients.

As with all statistical methods, propensity score matching comes with assumptions that must be understood to appropriately interpret the results presented. The stable unit treatment value assumption requires that the outcome of one subject is unaffected by the particular assignment of all other subjects. It seems reasonable to assume that the occurrence of hepatocellular carcinoma in one patient is not influenced by the group assignment of other patients.

Another facet of this assumption indicates that there is only 1 version of the treatment. Since patient treatment took place over 16 years and involved a variety of settings (a large university medical center, a community gastroenterology clinic, and community primary care centers), it is reasonable to consider whether, in this collection of patients, the unitary treatment assumption was appropriate.

Propensity score matching is also dependent on the positivity assumption, which states that all subjects in the analysis have some probability of receiving the treatment. To see whether this assumption holds, investigators can compare the distribution of propensity scores in the group receiving treatment to the distribution of scores in the control group. These distributions should be similar, for matching to be appropriate. In the study presented in this issue, all treated patients were matched with a reasonably tight caliper, suggesting that the positivity

Received 22 June 2018; editorial decision 22 June 2018; accepted 25 June 2018; published online July 5, 2018.

Correspondence: D. Kaplan, MD, PhD, Case Western Reserve University, Cleveland, OH (david.kaplan@case.edu).

**The Journal of Infectious Diseases® 2019;219:1–2**



DOI: 10.1093/infdis/jiy392

assumption was not violated. However, showing the distributions is preferable because it would enhance the readers' capacity to assess the analysis.

Another important assumption involves the inclusion of all significant covariates. In other words, the analysis is dependent on the model chosen by the investigators. The variables selected for the model include age, sex, HBV e antigen positivity, HBV DNA positivity, alanine aminotransferase level, baseline cirrhosis status, and follow-up time. These covariates seem reasonable and have produced a result that makes sense, but they are not the only values that may be significant.

More refined models would include molecular, immunological, and genomic covariates. For instance, various oncogenic pathways have been implicated in the pathogenesis of HBV-associated hepatocellular carcinoma [6]: telomerase reactivation, Wnt signaling, cell cycle, oxidative stress, epigenetics, the PI3K-Akt-mTor pathway, the MAPK pathway, JAK/Stat activation, and the retinoblastoma pathway. An analysis of these variables would certainly enrich the model. Of course, these data are not typically available in the medical record and would need to be collected specifically for the analysis.

Additionally, HBV genetic variants are not equivalently associated with the development of hepatocellular carcinoma [7]. Although there are conflicting results that make interpretation of the findings fuzzy, the genetic composition of the virus is certainly important in the development of cancer. For instance, certain mutations in the HBV X gene are likely to be involved in oncogenesis [6].

Because of the obvious sex disparity in the development of hepatocellular carcinoma, a consideration of sex hormone levels may be informative [6]. Sex may be a substitute for levels of sex hormones in the model, but at present we do not know which covariate better predicts the occurrence of hepatocellular carcinoma in HBV-infected persons.

As with other regressions, multicollinearity of the covariates must be avoided in logistic regression. It is likely that many of the molecular, immunological, and genomic factors that could be included in the model are reflected in the covariates that Nguyen et al used. Whether the predictive power of the model can be improved by including different covariates remains to be seen. However, the power of the interpretation of the results would certainly be enhanced by selecting specific parameters that imply significance in regard to diagnosis, patient management, and treatment.

Propensity score matching for causal inference in observational clinical studies is an important and powerful tool for physician-scientists. It allows for the use of data that would otherwise be difficult to interpret. In the study by Nguyen et al, the results that were obtained are not controversial and make sense. The real value of propensity score matching will be seen when this technology uncovers unexpected and controversial findings.

## Note

***Potential conflicts of interest.*** Author certifies no potential conflicts of interest. The author has submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

### References


1. Pearl J, Glymour M, Jewell NP. Causal inference in statistics: a primer. West Sussex, UK: Wiley, **2016**.
2. Nguyen MH, Yang H, Le A, et al. Reduced incidence of hepatocellular carcinoma with tenofovir in patients who have chronic hepatitis B with or without cirrhosis—a propensity score matched study. J Infect Dis **2019**; 219:10–8.
3. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika **1983**; 70:41–55.
4. Raynor WJ. Caliper pair-matching on a continuous variable in case-control studies. Commun Stat Theory Methods **1983**; 12:1499–509.
5. Lunt D. Selecting an appropriate caliper can be essential for achieving good balance with propensity score matching. Amer J Epidemiol **2014**; 179:226–35.
6. Xu W, Yu J, Wong VW. Mechanism and prediction of HCC development in HBV infection. Best Pract Res Clin Gastroenterol **2017**; 31:291–8.
7. Pollicino T, Saitta C, Raimondo G. Hepatocellular carcinoma: the point of view of the hepatitis B virus. Carcinogenesis **2011**; 32:1122–32.